Proceedings of the 2000 Winter Simulation Conference
J. A. Joines, R. R. Barton, K. Kang, and P. A. Fishwick, eds.

INTEGRATING DYNAMIC FAB CAPACITY AND AUTOMATION MODELS FOR 300MM SEMICONDUCTOR MANUFACTURING

Chad D. DeJong
Seth A. Fischbein

Intel Corporation
5000 West Chandler Boulevard
Mail Stop CH3-96
Chandler, AZ 85226, U.S.A.

ABSTRACT

Semiconductor fabrication facilities (fabs) continue to expand in both complexity and volume. As a result, integrated models are required to determine even high level impacts to key success indicators. In order to gain insight into how the components of a factory impact performance metrics, Intel uses an integrated discrete-event simulation modeling approach. Two models, one fab capacity and one automation model, are used. This paper discusses the methodology for building and integrating both models, and the results from using this method.

Both the fab capacity and automation models have a variety of input parameters that are required to drive the simulation. In addition, each model produces output parameters, some of which are used as inputs to the other model. An iterative feedback technique eventually results in a convergence on the appropriate data to feed the fab capacity model, which enables Intel to determine the impact of automation on 300mm wafer semiconductor manufacturing, and predict factory performance. Intel's approach provides the capability to use the models in stand-alone mode for specific fab-only or automation-only analyses, and also to take on any number of analyses via model communication. Intel continues to search for new applications for these merged models to answer strategic operational questions.

1 PROBLEM STATEMENT

Semiconductor fabs continue to expand in both complexity and volume. Correspondingly, the challenges of accurately modeling fab tools, labor, and material handling equipment continue to increase. Integrated, highly specific models are required to determine even high level impacts to units out, throughput time, WIP turns, and other key success indicators. This is a difficult undertaking, but is required to provide the level of decision support that Intel needs.

2 INTRODUCTION AND BACKGROUND

The focus of this paper is Intel's use of dynamic DES for two general types of models - fab capacity models and automation models. Both types of models assume a standard 300mm high-volume manufacturing (HVM) layout. Thus, model specifics such as the number of tools, labor requirements, material handling, automation, and layout (bays and aisles) are consistent in both models. Refer to Figure 1 for a graphical description.

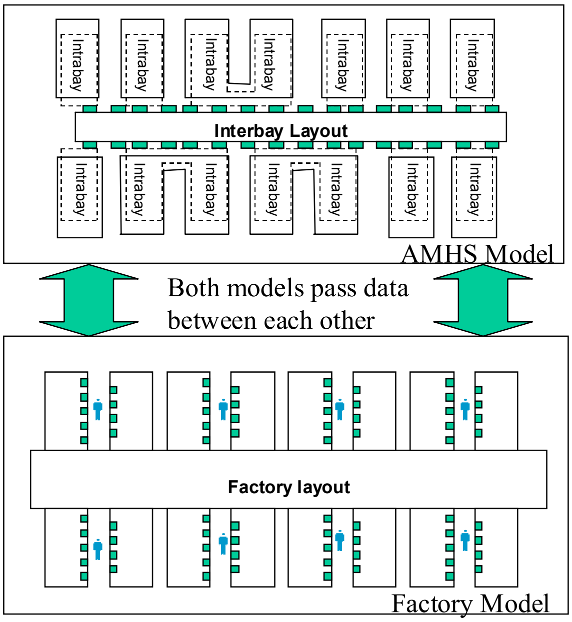

Figure 1: Integrated Capacity Model

The fab capacity model presents several modeling challenges, even without the AMHS / station location component. There are a number of peculiarities to semiconductor manufacturing that make DES modeling extremely complex. Two major characteristics of the fab process flow differentiate it from other industrial process flows. The first is its size. A full process route can contain up to five hundred individual steps. Second, the level of re-entrance in the process flow is extremely high. A lot may visit the same tool (or tool type) seven to eight times through the course of the process route.

Representing tool behavior in semiconductor fabrication is another challenge in building a capacity model. Tool dedication, queuing micro policies, and reticle (photolithographic mask) allocation must all be accounted for in the model. Another component of this behavior is in scheduled and unscheduled downtime events. To accurately model tool availability, vast amounts of data must be integrated in the form of preventative maintenance schedules, setup time requirements, and down (failure) events.

Representing the effects of labor is another significant challenge. Labor can have significant impact on fab performance, either through job certification, clustering (cross-training) strategies, or shift scheduling. While it is extremely difficult to model human behavior and decision making, this model estimates the delays and inefficiencies caused by human factors, and evaluates their impacts on the model's performance. An example of the complex decision making which human operators are responsible for is in lot scheduling. A facility will have multiple product types in production at any one time. Each product can have a different process route with different tool requirements, different dedication requirements, and different setups than the other products in the fab.

Automation modeling contains different but equally challenging complexity. Automation in semiconductor factories can be separated into three distinct classes – software, interbay hardware, and intrabay hardware. At a high level, automation software typically consists of a shop floor controls system, which exchanges information with the aforementioned AMHS and an automated equipment control system. Communication between the materials and equipment controls systems is critical to successfully moving lots between and within bays, as the lots follow their processing routes. The efficiency of the communication protocols, as well as the specific lot and stocker dispatch rules configured, have a great impact to total automation performance. Interbay systems are fab capacity wide systems which move material (lots) between bays or functional areas throughout the fab capacity, and are typically monorail-type movement systems. Intrabay systems are also monorail-type movement systems, but are local to a bay, or subset of bays. The interbay and intrabay systems interface through AS/RS machines (stockers), which consist of a robot and appropriate ports through which to transfer lots.

In order to gain insight of how the above components impact ultimate fab indicators (units out, WIP, TPT), and how they should be configured and used, Intel uses an integrated modeling approach. Specifically, this paper discusses the methodology for building and integrating both models, and the results from using this method.

3 MODELING METHODOLOGY

The fab capacity model has a variety of input parameters that are required to drive the simulation. It requires inputs of MHS delivery times to determine the delay time before a lot arrives in a tool family's queue to be available for processing, and how long it takes to deliver the lot from the queue to the tools load port. This information is provided by the automation model. The fab capacity model also uses loadport configuration information, such as how many loadports a tool has and how internal tool buffering of WIP occurs. In addition, the model requires that detailed labor clustering schemes be specified at the start of the simulation run. Labor can have a significant effect on fab capacity performance and as a result operators must be cross-trained appropriately and assigned adequately. The fab capacity model needs a highly detailed process flow, as each product type in a semiconductor fabrication facility can have its own particular route, processing time, tool dedication strategies, etc. The capacity model also has to have a destination table for push moves so the simulator knows where to send each lot to be processed, and how to flag a delivery request for the automation model.

The fab capacity model can track cycle time (throughput time), or the average time it takes for a lot to complete its route through the factory. Individual lot cycle time is tracked as well. Throughput time is a typical measure of a factory's performance, as it provides an indication of manufacturing agility. The model reports weekly WIP Turns, which is a measure of how quickly work is "turned over" on the floor. WIP turns are used to measure how efficiently a functional area or the factory is running. In addition, The fab capacity model records units (lots, wafers) out on average over various time windows, as well as on a cumulative basis. Finally, this model records tool starvation rates - time during which a tool was available to do work but no WIP was available to process (recorded as idle time) - and time during which a tool has selected a lot/batch to process but the lot/batch was unavailable (waiting for transportation). As part of the linkage to the automation model, The fab capacity model also has the ability to output detailed move requests in a manner that is readable by the automation model.

Automation modeling can also be defined in terms of its inputs and outputs. Inputs are as follows. The model needs a full fab layout (including tools and AMHS equipment for each bay), extensive interbay and intrabay vehicle scheduling logic, tool to bay and stocker to bay associations, move requirements from the fab capacity model, material control system and equipment control system software parameters, hardware specifications, (such as component cycle times) reliability metrics (for nodes, vehicles, and stockers), vehicle speed and acceleration, and placement of decision nodes. The outputs of the automation model are interbay wait and travel time for each loop, intrabay wait and travel time for each bay, vehicle statistics, target and achieved lot movement rates, and stocker and loadport utilization information. Refer to Figure 2 for a high level view of the fab capacity and automation model input and output structures.

The fab capacity model has two interfaces to the automation model - an input and an output interface. The input interface is a matrix that contains distributions for the interbay delivery times for a lot (i.e., stocker to stocker moves) and two arrays of distributions for intrabay delays (one each for delivery and retrieval). The output interface has the ability to activate move request reporting for a configurable length of time. This feature yields a “script” of moves to be read in by the automation model. This script generates a “traffic profile”, which is then handled appropriately. Like the fab capacity model, the automation model has an input and an output interface. The input interface is as described above.

The automation model reads in the script of move requests and generates lots accordingly. The output interface is the travel time taken by each move in the system. Post-processing generates fitted distributions from these times; these distributions populate the interbay delivery time matrix that is entered into the fab capacity model.

A protocol of this structure introduces a “chicken and egg” type of situation--one cannot generate accurate distributions without move scripts, and the move scripts cannot be generated without the distributions to drive travel time delays. As a result, the fab capacity model must first assume general distributions for the interbay and intrabay delivery time matrices as a starting point. From this, the fab capacity model generates a list of move requests, which is fed in to the automation model. In return, the automation model generates a new set of distributions, which are then provided to the fab capacity model. This iterative feedback technique eventually results in a convergence on the appropriate distributions to feed the fab capacity model, which then allows the analysis of the impact of automation on 300mm wafer semiconductor manufacturing.

4 RESULTS AND IMPLICATIONS

The key metrics or indicators for overall fab capacity performance are units out, WIP turns, and throughput time. The key metrics or indicators for automation performance are interbay and intrabay vehicle wait times and vehicle travel time (with a lot on board). Each simulated fab capacity and automation scenario reports these basic indicators. The results of these reports are analyzed from two perspectives. First, the authors analyzed the difference between baseline and final results. The amount of change between iterations and the number of iterations required for convergence are of particular interest. Convergence is defined as less than 2% difference in two sequential results. These results are represented in Figures 3 through 6.

The amounts of change between all iterations and especially between baseline and final results are of particular interest, because they provide insight to the level of impact automation has on final fab capacity performance, or vice-versa. For instance, because there is significant change in fab capacity performance between fab indicators for the baseline and the first iteration, then one can conclude that automation has an impact on fab performance (or that the baseline automation assumptions were wildly inaccurate). Conversely, if fab lot movement

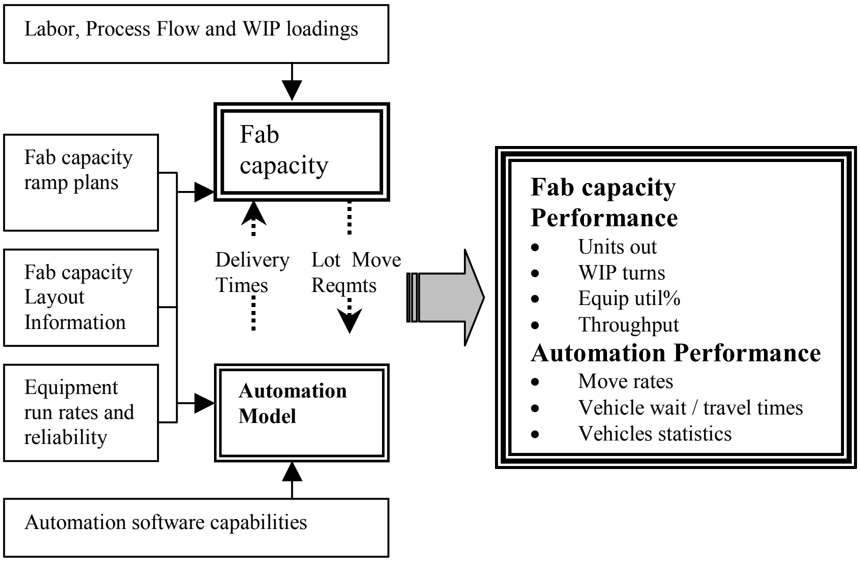

Figure 2: Model Input / Output Structure

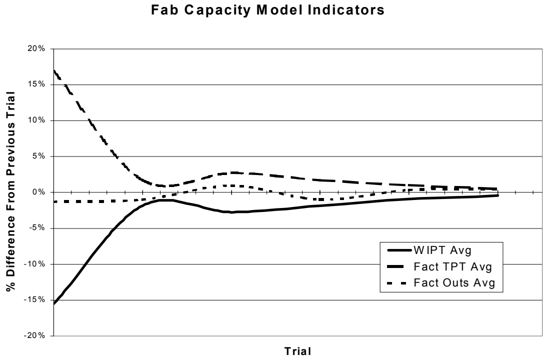

Figure 3: Fab Capacity Model Convergence Results

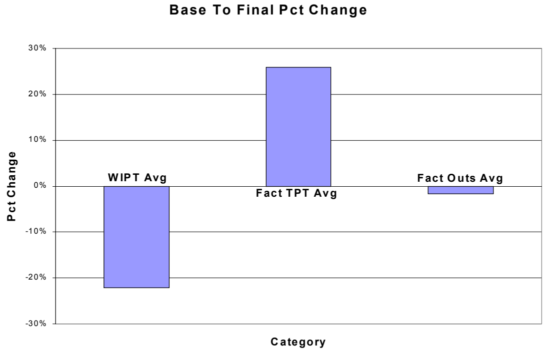

Figure 4: Percentage Difference between Fab Capacity Model Base Results and Final Results

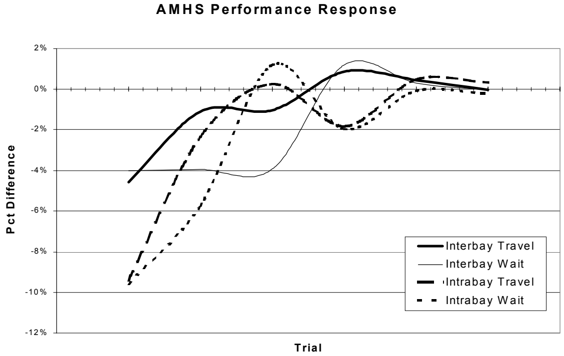

Figure 5: Automation Model Convergence Results

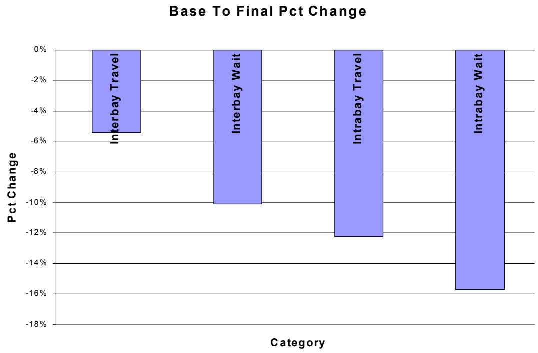

Figure 6: Percentage Difference between Automation Model Base Results and Final Results

requirements change between iterations, but automation performance has no significant change, one can conclude that the automation equipment is insensitive to these predicted changes in fab move requirements.

Figures 3 and 4 clearly indicate that factory performance converges on a stable equilibrium, in addition to demonstrating two manufacturing phenomena. The factory throughput time and WIP Turn curves are mirror images of each other due to the intimate relationship between cycle time, throughput, and WIP (Little's Law). The WIP turn metric uses factory inventory as the denominator, and hence is inversely proportional to cycle time. Also of note is that factory outs are not significantly impacted by changes in automation delays since the delivery times incurred by the AMHS do not constrain the system at any time.

Figure 5 clearly demonstrates that the automation model's metrics converge. The negative percent changes indicate that the automation model's performance improved over the assumed initial distributions and throughout the iterations.

Figure 6 provides evidence that as the moves requirement from the fab capacity model converges, both interbay and intrabay wait times are impacted more significantly than travel times. This shows that vehicle wait time is more sensitive to system loading than vehicle travel time.

5 CONCLUSIONS

Both fab capacity and DES models must be used to determine each system's impact to the other, and more importantly, the final predicted factory performance. Basic indicators such as units out, throughput time, WIP turns ultimately characterize fab performance. Intel has modeled the fab and automation models independently, but the models are now able to "communicate" with each other through passing of consistent data structures. Intel's method of passing data between the models, until fab and automation metrics have converged, has proven effective in predicting factory performance. This approach provides the capability to use the models in stand-alone mode if desired for specific fab-only or automation-only types of what-if analyses, and also to take on any number of analyses via integrated model communication. Intel continues to search for new applications for these merged models to answer strategic operational questions.

6 NEXT STEPS

The potential applications for this integrated model are far ranging. Thus, choosing the best follow-on focus areas will depend on support from customers within Intel. To date, there has been sparked interest in the following areas - preemptive downtimes, varying levels of automation, lot dispatching, and real-time model communication. Currently neither the fab capacity or automation model allows for the preemption of lots. Over a long period of time, statistically this has little (if any) significance. However, for greater accuracy and for easier simulation of short intervals, preemptive downtimes must eventually be enabled. Preemptive downtimes usually consist of "hard" failures that interrupt a tool's current task, force the lot(s) that was in process to return to the queue (usually with a high priority for re-selecting a station for work), and cause the tool to be unavailable for a period of time. However, with the introduction of loadport buffering, there arrives the problem of appropriately dealing with lots that are waiting in loadports to be processed. A later revision of the model will allow these lots to be "bumped" off of the loadports and back to the queue for the tool family.

At this time the models assume "lights-out" automation capabilities in the fab. That is, there is no operator interaction with any of the WIP on the floor in terms of handling or selection. Operators are used only for repairs and PM events. Future versions of the modeling approach will allow lower levels of automation to take place. For instance, operators may be needed to perform lot selection for the tools, or perform other tasks.

Currently, the fab capacity model performs lot selection from a station-centric viewpoint. That is, when a station finishes a lot or is informed that a lot has arrived in its family's queue, it performs its task selection routine to either select a lot or a batch of lots. With loadport buffering enabled, the station moves lots off of its family queue onto an available loadport. Lots that arrive at the family queue while stations are processing cause each station's task selection rule to run and attempt to select the lot(s) that arrived for placing on a loadport. Future versions of the fab capacity model will allow more intelligent dispatching of lots, such that work always arrives in the correct bay / stocker and get assigned to the correct tool.

Real-time model communication is challenging to accomplish in an efficient manner. The restriction of having to run the models on two different computers introduces a further "drag coefficient" communication over a network. In addition, the models' clocks and events lists must maintain synchronization, and finally the models must communicate the required information regarding lot departures and arrivals. However, the concepts are not difficult, and there are several applications in which this has previously been done. Intel plans to implement its own procedure in the near future.

ACKNOWLEDGMENTS

The authors collectively thank the following individuals, without whom this work would not have been possible: John Braunbeck, Dev Pillai, Madhav Rangaswami, and Ed Yellig for modeling concepts brainstorming and generation.

REFERENCES

AutoSimulations, Inc. 1999. *AutoMod User's Manual v 9.0*. Bountiful, Utah.

AutoSimulations, Inc. 1999. *AutoSched AP Customization Guide v 6.2*. Bountiful, Utah.

AutoSimulations, Inc. 1999. *AutoSched AP User's Guide v 6.2*. Bountiful, Utah.

Hopp, Wallace J. and Mark L. Spearman. 1996. *Factory Physics*. Chicago, Illinois: Richard D. Irwin.

Law, A.M. and W.D. Kelton. 1991. *Simulation Modeling & Analysis*, *Second Edition*. New York: McGraw-Hill, Inc.

AUTHOR BIOGRAPHIES

CHAD D. DEJONG is a Systems Engineer with the Operations Decision Support Technologies group at Intel Corporation. He is primarily responsible for the design, layout, and modeling of automated material handling systems. He earned a B.S. in Industrial and Operations Engineering from the University of Michigan. He also earned an M.S. in Industrial and Systems Engineering from Georgia Institute of Technology, and completed the Management of Technology certificate program. His previous professional background includes hospital and health care systems, and automotive component manufacturing. Current research and professional interests in the semiconductor industry are in whole factory capacity and operations modeling, supply chain simulation, model communication, model design and execution time reduction, and the statistical validation of modeling tools. His email address is `<chad.d.dejong@intel.com>`.

SETH A. FISCHBEIN is a Manufacturing Modeling Engineer with the Operational Decision Support Technologies group at Intel. His primary responsibilities consist of developing and sustaining dynamic capacity model logic for fab/sort manufacturing and assembly/test manufacturing. Seth received his B.S. in Operations Research and Industrial Engineering from Cornell University. Prior to his employment at Intel, Seth has worked as a manufacturing systems engineer for an Indonesian ceramics manufacturer and as a software developer for a Wall Street securities firm. Seth was also employed as a teaching assistant for simulation classes taught at Cornell in 1997. His current interests include dynamic semiconductor fab capacity modeling, interface development and design, and simulation engine development and model runtime reduction. Seth can be reached via email at `<seth.a.fischbein@intel.com>`.